Author(s):	van der Loos, Matthjis; Rietveld, Cornelius; Eklund, Nina; Kähönen, Mika; Lehtimäki, Terho **et al.**
Title:	The Molecular Genetic Architecture of Self-Employment
Year:	2013
Journal Title:	Plos ONE
Vol and number:	8 : 4
Pages:	1-15
ISSN:	1932-6203
Discipline:	Biomedicine
School /Other Unit:	School of Medicine
Item Type:	Journal Article
Language:	en
DOI:	http://dx.doi.org/10.1371/journal.pone.0060542
URN:	URN:NBN:fi:uta-201305161096
URL:	http://dx.doi.org/10.1371/journal.pone.0060542

OPEN ACCESS Freely available online

PLOS ONE

The Molecular Genetic Architecture of Self-Employment

Matthijs J. H. M. van der Loos[1,2]*, Cornelius A. Rietveld[1,2], Niina Eklund[3,4], Philipp D. Koellinger[1,2], Fernando Rivadeneira[2,5], Gonçalo R. Abecasis[6], Georgina A. Ankra-Badu[7], Sebastian E. Baumeister[8], Daniel J. Benjamin[9], Reiner Biffar[10], Stefan Blankenberg[11], Dorret I. Boomsma[12], David Cesarini[13], Francesco Cucca[14], Eco J. C. de Geus[12], George Dedoussis[15], Panos Deloukas[16], Maria Dimitriou[15], Guðny Eiriksdottir[17], Johan Eriksson[18,19,20,21,22], Christian Gieger[23], Vilmundur Gudnason[17,24], Birgit Höhne[23,25], Rolf Holle[26], Jouke-Jan Hottenga[12], Aaron Isaacs[27,28], Marjo-Riitta Järvelin[29,30,31], Magnus Johannesson[32], Marika Kaakinen[29], Mika Kähönen[33,34], Stavroula Kanoni[16], Maarit A. Laaksonen[35], Jari Lahti[36], Lenore J. Launer[37], Terho Lehtimäki[38,39], Marisa Loitfelder[40], Patrik K. E. Magnusson[41], Silvia Naitza[14], Ben A. Oostra[42], Markus Perola[3,4,43], Katja Petrovic[44], Lydia Quaye[7], Olli Raitakari[45,46], Samuli Ripatti[3,4,16], Paul Scheet[47], David Schlessinger[48], Carsten O. Schmidt[8], Helena Schmidt[49], Reinhold Schmidt[40], Andrea Senft[50], Albert V. Smith[17,24], Timothy D. Spector[7], Ida Surakka[3,4], Rauli Svento[51], Antonio Terracciano[48,52], Emmi Tikkanen[3,4], Cornelia M. van Duijn[27,28], Jorma Viikari[53,54], Henry Völzke[8], H. -Erich Wichmann[55,56,57], Philipp S. Wild[58,59], Sara M. Willems[27], Gonneke Willemsen[12], Frank J. A. van Rooij[2], Patrick J. F. Groenen[60], André G. Uitterlinden[2,5], Albert Hofman[2], A. Roy Thurik[1,61,62]

1 Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands, **2** Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, **3** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, **4** Public Health Genomics Unit, Department of Chronic Disease Prevention, The National Institute for Health and Welfare (THL), Helsinki, Finland, **5** Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands, **6** Center for Statistical Genetics, Department of Biostatistics, University of Michigan School of Public Health, University of Michigan Ann Arbor, Michigan, United States of America, **7** Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom, **8** Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, **9** Department of Economics, Cornell University, Ithaca, New York, United States of America, **10** Department of Prosthetic Dentistry, Gerodontology and Biomaterials, Centre of Oral Health, University of Greifswald, Greifswald, Germany, **11** Department of General and Interventional Cardiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, **12** Netherlands Twin Register, Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands, **13** Center for Experimental Social Science, Department of Economics, New York University, New York, New York, United States of America, **14** Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Cagliari, Italy, **15** Department of Nutrition and Dietetics, Harokopio University of Athens, Athens, Greece, **16** Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom, **17** Icelandic Heart Association, Kopavogur, Iceland, **18** Diabetes Prevention Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare (THL), Helsinki, Finland, **19** Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland, **20** Folkhälsan Research Center, Helsinki, Finland, **21** Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland, **22** Vaasa Central Hospital, Vaasa, Finland, **23** Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany, **24** Department of Medicine, University of Iceland, Reykjavik, Iceland, **25** Institute of Epidemiology II, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany, **26** Institute of Health Economics and Health Care Management, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany, **27** Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands, **28** Centre for Medical Systems Biology, Leiden, The Netherlands, **29** Institute of Health Sciences and Biocenter Oulu, University of Oulu, Oulu, Finland, **30** Department of Life Course and Services, National Institute for Health and Welfare, Oulu, Finland, **31** Department of Epidemiology and Biostatistics, MRC-HPA Centre for Environment and Health, Imperial College London, London, United Kingdom, **32** Department of Economics, Stockholm School of Economics, Stockholm, Sweden, **33** Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland, **34** Department of Clinical Physiology, University of Tampere School of Medicine, Tampere, Finland, **35** Population Health Research Unit, Department of Health, Functional Capacity and Welfare, National Institute for Health and Welfare (THL), Helsinki, Finland, **36** Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland, **37** Laboratory of Epidemiology, Demography, and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, United States of America, **38** Department of Clinical Chemistry, Fimlab Laboratories, Tampere University Hospital, Tampere, Finland, **39** Department of Clinical Chemistry, University of Tampere School of Medicine, Tampere, Finland, **40** Division for Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria, **41** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **42** Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands, **43** Estonian Genome Center, University of Tartu, Tartu, Estonia, **44** Division of General Neurology, Department of Neurology, General Hospital and Medical University of Graz, Graz, Austria, **45** Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland, **46** Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland, **47** Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **48** National Institute on Aging, National Institutes of Health, Baltimore, Maryland, United States of America, **49** Institute of Molecular Biology and Biochemistry, Medical University of Graz, Graz, Austria, **50** Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Lübeck, Germany, **51** Department of Economics, Oulu Business School, University of Oulu, Oulu, Finland, **52** College of Medicine, Florida State University, Tallahassee, Florida, United States of America, **53** Department of Medicine, Turku University Hospital, Turku, Finland, **54** Department of Medicine, University of Turku, Turku, Finland, **55** Institute of Epidemiology I, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany, **56** Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, **57** Klinikum Grosshadern, Munich, Germany, **58** Center for Thrombosis and Hemostasis, University Medical Center Mainz, Johannes Gutenberg University Mainz, Mainz, Germany, **59** Department of Medicine 2, University Medical Center Mainz, Johannes Gutenberg University Mainz, Mainz, Germany, **60** Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands, **61** Panteia, Zoetermeer, The Netherlands, **62** GSCM-Montpellier Business School, Montpellier, France

Abstract

Economic variables such as income, education, and occupation are known to affect mortality and morbidity, such as cardiovascular disease, and have also been shown to be partly heritable. However, very little is known about which genes influence economic variables, although these genes may have both a direct and an indirect effect on health. We report results from the first large-scale collaboration that studies the molecular genetic architecture of an economic variable–entrepreneurship–that was operationalized using self-employment, a widely-available proxy. Our results suggest that common SNPs when considered jointly explain about half of the narrow-sense heritability of self-employment estimated in twin data ($\sigma_g^2/\sigma_P^2 = 25\%$, $h^2 = 55\%$). However, a meta-analysis of genome-wide association studies across sixteen studies comprising 50,627 participants did not identify genome-wide significant SNPs. 58 SNPs with $p<10^{-5}$ were tested in a replication sample ($n = 3{,}271$), but none replicated. Furthermore, a gene-based test shows that none of the genes that were previously suggested in the literature to influence entrepreneurship reveal significant associations. Finally, SNP-based genetic scores that use results from the meta-analysis capture less than 0.2% of the variance in self-employment in an independent sample ($p \geq 0.039$). Our results are consistent with a highly polygenic molecular genetic architecture of self-employment, with many genetic variants of small effect. Although self-employment is a multi-faceted, heavily environmentally influenced, and biologically distal trait, our results are similar to those for other genetically complex and biologically more proximate outcomes, such as height, intelligence, personality, and several diseases.

Citation: van der Loos MJHM, Rietveld CA, Eklund N, Koellinger PD, Rivadeneira F, et al. (2013) The Molecular Genetic Architecture of Self-Employment. PLoS ONE 8(4): e60542. doi:10.1371/journal.pone.0060542

Editor: Stacey Cherny, University of Hong Kong, Hong Kong

Received December 12, 2012; **Accepted** February 27, 2013; **Published** April 4, 2013

Funding: AGES: The AGES–Reykjavik Study is funded by National Institutes of Health contract N01-AG-12100, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament); ASPS: The research reported in this article was funded by the Austrian Science Fond (FWF) grant number P20545-P05 and P13180. The Medical University of Graz supports the databank of the ASPS; ERF: The genotyping for the ERF study was supported by EUROSPAN (European Special Populations Research Network) and the European Commission FP6 STRP grant (018947; LSHG-CT-2006-01947). The ERF study was further supported by grants from the Netherlands Organization for Scientific Research, Erasmus MC, the Centre for Medical Systems Biology (CMSB) and the Netherlands Brain Foundation (Hersenstichting Nederland); GHS: This work/the Gutenberg Health Study is funded through the government of Rheinland-Pfalz (''Stiftung Rheinland Pfalz für Innovation'', contract number AZ 961-386261/733), the research programs ''Wissen schafft Zukunft'' and ''Schwerpunkt Vaskuläre Prävention'' of the Johannes Gutenberg-University of Mainz and its contract with Boehringer Ingelheim and Philips Medical Systems including an unrestricted grant for the Gutenberg Health Study; H2000: The study was funded mainly by the budgetary funds of National Institute for Health and Welfare (THL). The Finnish Centre for Pensions (ETK), the Social Insurance Institution of Finland (KELA), the Local Government Pensions Institution (KEVA) and other organisations listed on the website of the survey (http://www.terveys2000.fi) also contributed to funding; HBCS: The Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland (Grant No. 120315 and 129287 to EW, 1129457 and 1216965 to KR, 120386 and 125876 to JGE), the Finnish Diabetes Research Society, Folkhälsan Research Foundation, Novo Nordisk Foundation, Finska Läkaresällskapet, the European Science Foundation (EuroSTRESS), the Wellcome Trust (Grant No. 89061/Z/09/Z and 089062/Z/09/Z), Samfundet Folkhälsan, Finska Läkaresällskapet and the Signe and Ane Gyllenberg foundation. Markus Perola is partly financially supported for this work by the Finnish Academy SALVE program ''Pubgensense'' 129322; HRS: The Health and Retirement Study is sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan; KORA S4: The KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF). Part of this work was financed by the German National Genome Research Network (NGFN). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ; NFBC1966: Academy of Finland [project grants 104781, 120315, 129418 and Center of Excellence in Complex Disease Genetics and SALVE], University Hospital Oulu, Biocenter, University of Oulu, Finland (75617), the European Commission [EURO-BLCS, Framework 5 award QLG1-CT-2000-01643], NHLBI [5R01HL087679-02] through the STAMPEED program [1RL1MH083268-01], NIH/NIMH [5R01MH63706:02], ENGAGE project and grant agreement [HEALTH-F4-2007-201413], and the Medical Research Council, UK [G0500539, G0600705, PrevMetSyn/SALVE]; NTR: The Netherlands Twin Register data collection and genotyping has been funded by the Netherlands Organization for Scientific Research (NWO: MagW/ZonMW grants 904-61-090, 985-10-002,904-61-193,480-04-004, 400-05-717, Addiction-31160008 Middelgroot-911-09-032, Spinozapremie 56-464-14192), Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI–NL, 184.021.007), the VU University's Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA), the European Science Foundation (ESF, EU/QLRT-2001-01254), the European Community's Seventh Framework Program (FP7/2007-2013), ENGAGE (HEALTH-F4-2007-201413); the European Science Council (ERC Advanced, 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, the (NIMH, MH081802) and by the Grand Opportunity grants 1RC2MH089951-01 and 1RC2 MH089995-01 from the NIMH; RS: The Rotterdam Study is funded by the Erasmus Medical Center and Erasmus University, Rotterdam, the Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The GWAS database of the Rotterdam Study was funded by the Netherlands Organisation for Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), and the Netherlands Genomics Initiative (NGI)/Netherlands Consortium for Healthy Aging (NCHA) project nr. 050-060-810. Statistical analyses were partly carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam; SardiNIA: This research was supported in part by the Intramural Research Program of the National Institute on Aging, NIH, and by the National Institute on Aging contract NO1-AG-1-2109 to the SardiNIA/ProgeNIA team; SHIP: SHIP is part of the Community Medicine Research net (www.community-medicine.de) and the Greifswald Approach to Individualized Medicine (GANI_MED) consortium (www.gani-med.de) of the University Medicine Greifswald, Germany, which are funded by the Federal Ministry of Education and Research (BMBF 01ZZ9603, 01ZZ0103 and 03IS2061A), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the InterSystems GmbH; STR: Financial support was received from The Swedish Council for Working Life and Social Research, The Jan Wallander and Tom Hedelius Foundation, the Ragnar Söderberg Foundation, the Ministry for Higher Education, the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF), and the Heart and Lung foundation no. 20070481; THISEAS: Recruitment for THISEAS was partially funded by a research grant (PENED 2003) from the Greek General Secretary of Research and Technology. Genotyping was supported by the Wellcome Trust Sanger Institute; TwinsUK: The study was funded by the Wellcome Trust, the European Community's Seventh Framework Programme (FP7/2007-2013), and the ENGAGE project grant agreement (HEALTH-F4-2007-201413). The study also receives support from the Dept of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. TDS is an NIHR senior Investigator and is holder of an ERC Advanced Principal Investigator award. Genotyping was performed by The Wellcome

Trust Sanger Institute, support of the National Eye Institute via an NIH/CIDR genotyping project; YFS: The Young Finns Study has been financially supported by the Academy of Finland: grants 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi),and 41071 (Skidi), the Social Insurance Institution of Finland, Kuopio, Tampere and Turku University Hospital Medical Funds (grant 9M048 for TeLeht), Juho Vainio Foundation, Paavo Nurmi Foundation, Finnish Foundation of Cardiovascular Research and Finnish Cultural Foundation, Tampere Tuberculosis Foundation and Emil Aaltonen Foundation (TL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors would like to declare financial support by Philips Medical Systems for the Gutenberg Health Study (GHS). Philips Medical Systems provided the ultrasound machines for the GHS and had no role in the current research. Hence, funding by Philips Medical Systems does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: mvanderloos@ese.eur.nl

Introduction

Economic variables such as income, education, and occupation are well-known to be related to health outcomes and longevity [1–10]. Specifically, there is a consistent inverse relation between indicators of socioeconomic status and cardiovascular disease [11]. For example, occupational choice is associated with the incidence of coronary heart disease among women [12]. Intriguingly, health outcomes, longevity, income, educational attainment, and occupational choice have all been shown to be partly heritable (see ref. [13] for complex diseases, refs. [14–17] for longevity, refs. [18–22] for education, refs. [23–25] for income, and refs. [26–28] for occupational choice). This suggests that the same genetic factors could be linked to socioeconomic status and health outcomes, or that indirect causal pathways from genetic variants to health outcomes exist that are mediated by individual behavior and the environment. For example, a potential mismatch between personal disposition and occupational choice may result in stress and decreased happiness, which have been shown to negatively affect (cardiovascular) disease incidence and longevity [29–32]. Therefore, knowledge about the specific molecular genetic architecture of socioeconomic variables and about the effects of mismatches between genetic predispositions and realized choices could yield important insights for epidemiology and public health policy. Unfortunately, most efforts to investigate the influence of genes on economic variables were until now limited to candidate gene studies that often failed to replicate later [33,34].

This study reports results from the first large-scale collaboration that studies the molecular genetic architecture of a specific economic behavior–entrepreneurship–using data from high-density SNP arrays. Entrepreneurship has been associated with poor health [35], increased stress [36], relatively low average incomes [37], but also with greater job and life satisfaction [38–40]. The analysis of entrepreneurship is complicated by the fact that it is a multi-faceted phenomenon [41]. Individuals may engage in entrepreneurial activity for a variety of reasons. For example, certain individuals may be motivated to pursue a business opportunity or to gain independence, whereas others may do so because of unemployment and a lack of viable alternatives in paid employment. Despite this complexity, empirical evidence suggests that entrepreneurship tends to run in families [42–47], and recent twin studies consistently estimate the heritability of this behavior to be on the order of 50% [26–28]. As these results suggest that entrepreneurship is partly influenced by genetic variation, specific markers that are associated with entrepreneurship should, in principle, exist. Research that is aimed at discovering these specific markers has thus far been limited to one candidate gene study. This study [48] found evidence for an association between a specific genetic variant in the *DRD3* gene and entrepreneurship in a sample of $n = 1{,}335$. However, a more recent study [49] failed to replicate this association in three larger samples of $n = 5{,}374$, $n = 2{,}066$, and $n = 1{,}925$.

The molecular genetic architecture of entrepreneurship therefore remains largely unknown. A variety of alternative architectures could account for heritable variation. For example, there may be a small number of rare variants with strong effects, multiple common variants with small or modest effects, or some combination of these possibilities [50,51]. Therefore, we aimed to identify the molecular genetic architecture of entrepreneurship to facilitate a more sophisticated understanding of the nature of the associated heritable variation.

We use self-employment as a proxy for entrepreneurship in this study, which is the most widely available proxy for entrepreneurship. Self-employment is defined as having started, owned, and managed a business. Initially, we used a classical twin design to estimate the heritability of the tendency to engage in self-employment. We performed this analysis to determine the comparability of our results with (1) estimates of previous twin studies, and (2) estimates from a novel method from molecular genetics. This recently described method [52] is used here to quantify the proportion of variance that is explained by common SNPs (and unknown causal variants that are in linkage disequilibrium with these SNPs) in the tendency to engage in self-employment.

Furthermore, we performed a meta-analysis of genome-wide association studies (GWASs) of self-employment from sixteen studies to identify genetic variants that are robustly associated with self-employment. Together, these studies comprised 50,627 participants of European ancestry who are part of the Gentrepreneur Consortium [53,54]. This study is the first large-scale effort to identify common genetic variants that are associated with an economic variable. We also tested whether self-employment could be predicted out-of-sample solely using genotype data and the results of our meta-analysis.

Theoretical and empirical evidence from entrepreneurship research suggests that there may be differences between males and females with respect to the type of businesses they start. These differences also extend to individuals' motivations, goals, and resources [55–59] and exist because women face different–and typically more–barriers to entrepreneurship than men [60–62]. Therefore, we performed both pooled and sex-stratified analyses for all of our investigations.

Materials and Methods

Participating studies and self-employment measures

The analyses were performed within the Gentrepreneur Consortium [53,54], which included two out of the five studies that participate in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium [63] and fourteen additional studies. The discovery studies included the Age, Gene/Environment Susceptibility–Reykjavik Study (AGES), the Austrian Stroke Prevention Study (ASPS), the Erasmus Rucphen Family study (ERF), the Gutenberg Health Study (GHS), Health 2000 (H2000), the Helsinki Birth Cohort Study (HBCS), the Health and Retirement Study (HRS), the Cooperative Health Research in the Region of Augsburg (KORA S4), the Northern Finland Birth Cohort 1966 (NFBC1966), the Netherlands Twin

Register Cohort 1 (NTR1), the Netherlands Twin Register Cohort 2 (NTR2), the Rotterdam Study Baseline (RS-I), the Rotterdam Study Extension of Baseline (RS-II), the Rotterdam Study Young (RS-III), the SardiNIA Study of Aging (SardiNIA), the Study of Health in Pomerania (SHIP), The Hellenic study of Interactions between SNPs & Eating in Atherosclerosis Susceptibility (THISEAS), the UK Adult Twin Registry (TwinsUK), and the Cardiovascular Risk in Young Finns Study (YFS). The Swedish Twin Registry (STR) served as an *in silico* replication study, as genome-wide data were only available following the completion of the discovery stage.

The studies collected data regarding occupational status using questionnaires or interviews, from which self-employment status was distilled. Self-employment measures were defined in collaboration with the consortium leaders to minimize heterogeneity across participating studies. The cases were defined as individuals who were self-employed at least once, and the controls were defined as individuals who were never self-employed during their working life. However, for a number of studies, reliable data regarding work-life history were unavailable, possibly resulting in the inclusion of previously self-employed individuals in the control group. The details regarding the background and self-employment measures of each of the discovery studies and of the replication study are given in Table S1.

Ethics statement

All participating studies were approved by the relevant institutional review boards or the local research ethics committees, including the Icelandic National Bioethics Committee (VSN: 00-063), the Icelandic Data Protection Authority, and the Institutional Review Board for the National Institute on Aging (AGES); the Ethics Committee of the Medical Faculty of the University of Graz (ASPS); the Medical Ethics Committee at Erasmus University which approved the protocols for the ascertainment and examination of human subjects (ERF); the local ethics committee and data safety commissioner, the sampling design was approved by the federal data safety commissioner (GHS); the Ethics Committee for Epidemiology and Public Health in the Hospital District of Helsinki and Uusimaa in Finland, in accordance with the ethical standards of the Declaration of Helsinki (H2000); the Ethics Committee of Epidemiology and Public Health of the Hospital District of Helsinki and Uusimaa (HBCS); the Health Sciences Institutional Review Board at the University of Michigan (HRS); the Ethics Committee of the Bavarian Medical Association (KORA S4); the Ethics Committee of the University Hospital of Oulu (NFBC1966); the VU University Medical Ethical Committee (NTR); the Medical Ethics Committee of the Erasmus Medical Center (RS); the local Ethics Committee for the Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche and the MedStar Research Institute, responsible for intramural research at the National Institute of Aging (SardiNIA); the Ethics Committee of the University of Greifswald (SHIP), the Ethical Review Board in Stockholm (STR); the Bioethics Committee of the Harokopio University of Athens (THISEAS); the NRES Committee London-Westminster (TwinsUK); the local Ethics Committees of the participating universities (YFS). Written informed consent was provided by all of the participants.

Genotyping, imputation, and quality control

The seventeen participating studies used a variety of commercially available SNP genotyping platforms to genotype their participants. Each study performed quality control of their genotypic data and imputed the genotypes of each participant to a common set of approximately 2.5 million SNPs from the HapMap CEU population. The exceptions to this were THISEAS, which only supplied results for directly genotyped SNPs, and HRS, which imputed to the 1,000 Genomes Project Phase I v3 panel. Prior to the meta-analysis, we performed parallel quality control of the association results for each study. SNPs were excluded on the basis of minor allele frequency (MAF$<$0.01 or MAF$<$0.05 if deemed necessary) and if the imputation quality (a measure of the observed variance divided by the expected variance of the imputed allele dosage from the imputation software output) was less than 0.4. Following these exclusions, approximately 2.4 million SNPs remained. Study-specific details regarding the genotyping, imputation, and quality control are given in Table S2.

Statistical analysis

Tetrachoric correlations were used to calculate self-employment correlations for MZ and DZ twin pairs. This analysis assumes a latent normally distributed tendency to engage in self-employment. We estimated the heritability of the tendency to engage in self-employment in the replication study using standard twin study methods, which were implemented in the program Mx [64]. Only complete twin pairs with data regarding self-employment status were included in the analysis and opposite-sex DZ twin pairs were excluded, resulting in a final sample size of 4,464 individuals. Specifically, for pooled males and females, males only, and females only, we fitted the three following nested models using the maximum likelihood approach on the raw data: (1) a model including an additive genetic effect, a shared common environment effect, and an individual-specific environment effect (the *ACE* model); (2) a model that included only an additive genetic and an individual-specific environment effect (the *AE* model); and (3) a model including only a common environment effect and an individual-specific environment effect (the *CE* model). For all of the samples, we controlled for a z-score of age by estimating age-specific thresholds. For the pooled sample, we additionally controlled for sex in a similar way.

We used the method that was recently developed by Yang et al. [52] to estimate the proportion of variance in the tendency to engage in self-employment that is explained by all of the common genotyped SNPs. The method is implemented in the GCTA software [65] and hinges on the assumption that in a sample of unrelated individuals, environmental factors segregate independently in the pedigree from the degree of genetic relatedness. In contrast to the twin study design, genetic relatedness is not inferred from the pedigree but is estimated directly from genome-wide SNP data. Under the assumption of no confounding by environmental variables, we can then estimate the accounted-for variance by relating the estimated genetic relatedness between pairs of individuals to their phenotypic correlation. The resulting estimate is actually a lower bound of the heritability that is estimated from classic twin and family studies. The reason for this is that twin and family studies capture the variation that is due to all of the additive causal variants, whereas the more recently developed method only captures the variants that are either directly genotyped or in linkage disequilibrium.

We used a combined sample of individuals from one of the discovery studies (RS-I) and the replication study (STR) to estimate the accounted-for variance. We restricted the sample from each study to individuals for whom data regarding self-employment were available. Additionally, we included only one randomly selected individual from each family in the STR sample. A second round of quality control of the genotypic data was then performed for both studies. In the RS-I sample, we excluded 3,748 SNPs because they failed a test of Hardy-Weinberg equilibrium at

$p<1\times10^{-6}$. We removed 24,993 SNPs with minor allele frequencies that were lower than 0.01 and another 6,665 due to data missingness greater than 5%. In total, 5,374 individuals and 561,466 autosomal SNPs were included in the analysis. In the STR sample, we removed two SNPs because they failed a test of Hardy-Weinberg equilibrium at $p<1\times10^{-6}$. Another 628 SNPs with a minor allele frequency lower than 0.01 were removed, as were two SNPs with data missingness greater than 5%. Therefore, 643,924 autosomal SNPs and 2,589 individuals were included in the analysis.

We then estimated the genetic relationships among 7,963 individuals in the combined sample from the 301,115 common autosomal SNPs. We dropped one of any pair of individuals with an estimated genetic relationship that was >0.025 while maximizing the remaining sample size to exclude the possibility of ascribing shared environmental effects to genetic effects and/or including the effects of causal variants not correlated with the genotyped SNPs but captured by the pedigree. The maximum relatedness in the remaining sample of 6,223 individuals therefore approximately corresponds to cousins two to three times removed [52].

Next, the linear mixed model $y = \mu+g+e$ was fitted, where y is the binary phenotype, g the total additive genetic effect of the SNPs, and e is a residual effect. The restricted maximum likelihood (REML) was used to estimate the variance of the total additive genetic effect σ_g^2 of the SNPs by fitting the genetic relationships as the covariance structure. Because the analyzed phenotype is binary, σ_g^2 is the variance of the total additive genetics effects on the observed 0–1 scale. A latent normally distributed tendency to engage in self-employment was assumed when transforming the explained variance from the observed 0–1 scale to the latent scale using the transformation that is derived in the appendix of Dempster and Lerner [66]. For all of the analyses, we controlled for a z-score of age, study, and the first ten principal components of the genetic relationships of the combined sample. In the pooled sample, we also controlled for sex.

In addition to the Yang et al. [52] method, we employed a novel method developed by So et al. [67] that serves the same purpose, i.e., estimating the proportion of variance in the tendency to engage in self-employment that is explained by all of the common SNPs. However, in contrast to the Yang et al. [52] method, So et al.'s method does not require raw genotype data but attempts to recover the accounted-for variance from the meta-analysis results. Using PLINK [68], we restricted the meta-analysis results to SNPs that were present in the HapMap Phase II CEU panel (release 23a) and pruned those in strong linkage disequilibrium with other SNPs using a pairwise r^2 threshold of 0.25 in a window of 100 SNPs that slides in 25 SNP increments. After this procedure 172,742, 175,970, and 172,989 SNPs remained in the pooled males and females, males only, and females only sample, respectively. We used the Gaussian Kernel function, considered under the null-hypothesis of no association, and ran the simulation 500 times in each sample.

The genome-wide association analysis of self-employment was independently performed by each study according to a predefined analysis plan. The analyses were performed for pooled males and females, males only, and females only using an additive genetic model, controlling for age (≤29 [reference]; 30–39; 40–49;≥50) and sex in the pooled sample. To control for population stratification, the first four principal components of the genotypic data were also included if available. We provide details regarding the statistical analysis within each study in Table S2.

Following the association analyses, the genomic inflation factor λ was calculated for each sample to quantify any remaining population stratification or cryptic relatedness. The lowest inflation factor was 0.989, and the highest was 1.156, although this latter value was for a study that did not include the first four principal components of the genotypic data in the analysis (Table S3). Genomic control [69] was applied in samples with inflation factors that were greater than one by adjusting the test statistics.

We next performed fixed-effect meta-analyses of the association results from the discovery studies for pooled males and females, males only, and females only using METAL software [70]. Although the phenotype was defined as self-employment in each participating study, we could not harmonize the exact wording of the question on which the self-employment measure was based. In addition, the connotations of self-employment may depend to some extent on the level of economic development and culture. This may lead to unobserved gene-environment interactions that could introduce additional noise in the GWAS results pooled across studies. We combined the association results using weighted z-scores that were based on the p-values and the direction of the effects. This method first computes a per-study signed z-score for each SNP based on its p-value and the effect direction. The z-scores are then summed with weights that are proportional to the square root of the sample size of each study. Following the meta-analyses, only autosomal SNPs that were present in the Hapmap Phase II CEU panel (release 22, NCBI build 36) and in at least half of the contributing samples in each meta-analysis were retained prior to both reporting p-values and the creation of the Q–Q and Manhattan plots. We *a priori* set the genome-wide significance threshold to $p<5\times10^{-8}$. SNPs with $p<1\times10^{-5}$ were considered suggestive and also carried forward to the replication stage. The heterogeneity of the test statistics between the studies was assessed using the I^2 metric [71,72] and Cochran's Q statistic [73].

Replication was attempted for significant and suggestive SNPs from each meta-analysis using an *in silico* replication study comprising 3,271 individuals. The association results for these SNPs were looked up in the replication study and meta-analyzed together with the discovery samples for pooled males and females, males only, and females only. To adjust for family relationships in the replication study, we performed family-based association tests implemented in the MERLIN software [74].

We used the discovery meta-analyses results to calculate gene-based p-values using the VEGAS program [75]. The positions of the UCSC Genome Browser hg18 assembly were employed to assign SNPs to genes, which included regions that were ±50 kb from the 5′ and 3′ UTRs.

For the prediction analyses, we followed the approach that was pioneered by The International Schizophrenia Consortium [76] and used the association results from the discovery meta-analyses to predict self-employment in the STR. Specifically, twelve overlapping sets of SNPs that were nominally associated in the discovery meta-analyses were created for different significance thresholds ($p_T<0.01$, $p_T<0.05$, $p_T<0.1$, $p_T<0.2$, $p_T<0.3$, $p_T<0.4$, $p_T<0.5$, $p_T<0.6$, $p_T<0.7$, $p_T<0.8$, $p_T<0.9$, and $p_T\leq1$). These sets were used as inputs for score calculation in the STR. We restricted the STR sample to individuals for whom data regarding self-employment were available and included only one randomly selected individual from each family, resulting in a final sample size of 2,589 individuals for the prediction analyses.

Prior to calculating the scores for each individual in the STR, we followed [76] and selected all of the autosomal SNPs, pruning those in strong linkage disequilibrium with other SNPs. This process was performed using a pairwise r^2 threshold of 0.25 in a window of 200 SNPs that slides in five SNP increments. Following this exclusion process, 135,823 SNPs remained. The PLINK [68] 'score' function was then used to calculate the total score for each

individual in the STR. The score is defined as the sum of the number of score alleles, weighted by the estimated coefficients from the discovery meta-analyses, divided by the number of non-missing genotypes. If an individual was missing a genotype, it was imputed as the mean genotype based on the score allele frequency in the STR. On average, the score was calculated from approximately 120,000 SNPs given that (1) the coefficients were only estimated for SNPs in the HapMap CEU population in the discovery meta-analyses, and (2) the overlap with the genotyped SNPs was not perfect. Lastly, we regressed self-employment onto the score using a logistic regression model. The variance that was explained by the score was estimated using the Nagelkerke pseudo-R^2 of the fitted model. We also calculated the area under the receiver operating characteristic curve (AUC) to evaluate the prediction accuracy.

Results

Heritability of self-employment and the degree of variance that is accounted for by common SNPs

We used data from the Swedish Twin Registry (STR) and the classical twin design to estimate the heritability of the tendency to engage in self-employment. We computed the tetrachoric correlations between the tendencies to engage in self-employment within monozygotic (MZ) and dizygotic (DZ) twin pairs. Table 1 indicates that the correlations within the MZ twin pairs were consistently higher than within the DZ twin pairs for males only, for females only, and for pooled males and females. We note that the correlation within DZ twin pairs in the pooled sample was higher than for the DZ correlations in males and females when the two sexes are considered separately. This effect most likely results from imprecise estimation of the tetrachoric correlations due to the small number of cases. When we computed Pearson correlations, the pooled DZ twin pairs correlation was in between the male and female DZ twin pairs correlations. Applying Falconer's formula [77] to the correlations in Table 1, yields h^2 estimates of 0.39 for pooled males and females, 0.69 for males only, and 0.34 for females only.

A maximum likelihood approach was employed to estimate the relative contributions of the additive genetic (*A*), shared common environment (*C*), and individual-specific environment (*E*) components. This approach was performed using an *ACE* model and two nested submodels for pooled males and females, males only, and females only. Table 2 gives the estimates of the *A* component as 0.54 for pooled males and females, 0.67 for males only, and 0.38 for females only. The estimates of the *C* component were 0.01 for pooled males and females, 0.00 for males only, and 0.02 for females only. The *A* component was significant at the 95% confidence level for pooled males and females, and for males only, although the confidence intervals were very wide. This component was not significant for the females only analysis. However, the χ^2 test for goodness-of-fit and Akaike information criterion indicated that the *AE* model was the best-fitting model in all samples. In this submodel, the estimate for the *A* component for females only did not change markedly compared to the *ACE* model but was significant at the 95% confidence level. The estimates of the *A* component for pooled males and females, and males only were 0.55 and 0.67, respectively; these results were significant.

Table 1. Tetrachoric correlations in the tendency to engage in self-employment for MZ and DZ twin pairs in STR for pooled males and females, males only, and females only.

	Pooled		Males		Females	
	MZ	DZ	MZ	DZ	MZ	DZ
n	1,062	1,170	419	469	643	701
Concordant pairs	839	868	320	307	519	561
Discordant pairs	223	302	99	162	124	140
Pairwise concordance (%)	79.0	74.2	76.4	65.5	80.7	80.0
Tetrachoric ρ	0.560	0.363	0.677	0.332	0.401	0.230
s.e.	0.042	0.052	0.053	0.072	0.078	0.090

n refers to the number of twin pairs; s.e.: standard error.

The recently developed method by Yang et al. [52] was employed to estimate the degree of variance in the tendency to engage in self-employment that is explained by all of the genotyped autosomal SNPs in the GWAS datasets. The proportion of the explained variance was estimated for pooled males and females, males only, and females only. To maximize the power of the analysis, we used a combined sample of one of the discovery studies (Rotterdam Study Baseline [RS-I]) and the STR. We estimated that 25% ($p = 0.032$) of the variance in the tendency to engage in self-employment could be explained by the common genotyped autosomal SNPs for pooled males and females (Table 3). The variance that could be explained for males only and for females only was 25% ($p = 0.152$) and 0% ($p = 0.499$), respectively. The estimates for males and females separately were not significantly different from one other. The fact that the variance that is explained was zero for females is most likely due to the very low number of female cases ($n = 353$) compared to the number of controls ($n = 3{,}482$). The estimation of the explained variance is therefore very imprecise. We also estimated the variance that was explained for pooled males and females, males only, and females only in the RS-I and the STR separately. The estimates were not significant because the standard errors of these estimates depend heavily on the sample size. However, considered in their entirety, the results were consistent with the estimates that we present for the combined RS-I and STR samples. Overall, the results for pooled males and females and for males indicated that the degree of variance in the tendency to engage in self-employment that is explained by all of the common autosomal SNPs simultaneously is only approximately half of the narrow-sense heritability that is estimated using the STR and the classical twin design. Furthermore, estimates using the method developed by So et al. [67] also provide non-zero estimates for heritability. Specifically, the accounted-for variance was 7% for pooled males and females, 21% for males only, and 15% for females only. However, confidence intervals and standard errors could not be calculated for these estimates because not all raw genotype data were available, prohibiting further interpretation of these results.

Meta-analyses of genome-wide association studies

We performed genome-wide association analyses of self-employment using the data from sixteen discovery studies. These studies comprised 7,734 participants who had been self-employed at least once and 42,893 participants who did not report being self-employed. Table 4 includes the descriptive statistics for the studies. The mean ages in the pooled samples of males and females ranged from 31 to 68.8 years, and the average age across all of the studies was 53.4 years. Following independent association analyses for each study, we performed a fixed-effect meta-analysis of the study-level results for approximately 2.4 million SNPs using a pooled *z*-score approach.

The discovery meta-analysis Q–Q plot (Figure 1A) did not indicate a strong deviation for the lowest *p*-values. However, no

Table 2. Results of fitting *ACE*, *AE*, and *CE* models to the tendency to engage in self-employment in STR for pooled males and females, males only, and females only.

Sample	Model	*A*	(95% CI)	*C*	(95% CI)	*E*	(95% CI)	χ^2	*p*-value	AIC
Pooled	*ACE*	0.54	(0.25–0.63)	0.01	(0.00–0.25)	0.45	(0.37–0.55)	–	–	−4,707.96
	AE	0.55	(0.46–0.63)	–	–	0.45	(0.37–0.54)	0.01	0.929	−4,709.95
	CE	–	–	0.42	(0.35–0.49)	0.58	(0.51–0.65)	13.60	<0.001	−4,696.36
Males	*ACE*	0.67	(0.33–0.76)	0.00	(0.00–0.28)	0.33	(0.24–0.44)	–	–	−1,417.15
	AE	0.67	(0.56–0.76)	–	–	0.33	(0.24–0.44)	0.00	1.000	−1,419.15
	CE	–	–	0.50	(0.41–0.59)	0.50	(0.41–0.59)	14.27	<0.001	−1,404.88
Females	*ACE*	0.38	(0.00–0.53)	0.02	(0.00–0.38)	0.60	(0.47–0.76)	–	–	−3,276.62
	AE	0.40	(0.26–0.53)	–	–	0.60	(0.47–0.75)	0.01	0.919	−3,278.61
	CE	–	–	0.31	(0.19–0.42)	0.69	(0.58–0.81)	2.50	0.114	−3,276.12

For pooled males and females the analyses are based on 2,232 twin pairs (1,062 MZ and 1,170 DZ), for males only on 888 twin pairs (419 MZ and 469 DZ), and for females only on 1,344 twin pairs (643 MZ and 701 DZ). The share of self-employed was 21% for the pooled, 32% for the male, and 13% for the female sample. In all samples we controlled for age and in the pooled sample for sex; *A*: additive genetic component; *C*: shared common environment component; *E*: individual-specific environment component; 95% CI: 95% confidence interval; χ^2: χ^2 test for goodness-of-fit, the baseline model is the *ACE* model; AIC: Akaike information criterion.
doi:10.1371/journal.pone.0060542.t002

confounding issues related to population stratification, cryptic relatedness, or genotyping errors were detected, as no systematic deviation from the expectation under the null hypothesis of no association was observed [78]. As illustrated in the Manhattan plot (Figure 2A), we observed twenty SNPs with $4.1\times10^{-6}\leq p<1\times10^{-5}$ (Tables 5 and S4). The SNP with the lowest *p*-value, rs6906622 ($p=4.10\times10^{-6}$), was located near the *RNF144B* gene, with most studies indicating that the minor allele increased the probability of being self-employed (Table 5).

We next attempted to replicate *in silico* the twenty suggestive SNPs in the STR ($n=3{,}271$). Two of the twenty SNPs associated with self-employment were statistically significant at the 5% level in the replication study. However, the SNP effects were not in the same direction as in the majority of the discovery studies (Table S4), indicating that these SNPs were potential false positives. We then performed a combined meta-analysis of the discovery and replication studies. For all SNPs, the *p*-values were larger in the combined sample than in the discovery sample and did not reach genome-wide significance (Table S4).

The Q–Q plot for the male only meta-analysis (Figure 1B) gave a certain degree of suggestive evidence of association; however, no evidence of population stratification, cryptic relatedness, or genotyping errors was observed, as only certain SNPs–those with particularly low *p*-values–deviated from their expectation under the null hypothesis of no association. The female only meta-analysis Q–Q plot (Figure 1C) did not indicate a strong deviation for the lowest *p*-values and no evidence of population stratification, cryptic relatedness, or genotyping errors was observed. No SNPs reached genome-wide significance in the sex-stratified meta-analyses (Table 5), as can be observed in the Manhattan plots (Figures 2B and C). The male meta-analysis resulted in 22 suggestive SNPs with $p<1\times10^{-5}$, and the female meta-analysis resulted in sixteen suggestive SNPs (Tables 5, S5, and S6). The top SNP in males, rs6738407 ($p=1.52\times10^{-7}$), was located in the *HECW2* gene, and most studies reported that carrying the minor allele decreased the probability of being self-employed. The top SNP in females, rs2331548 ($p=1.93\times10^{-6}$), was located near the *CBR4* gene, and most studies estimated that carrying the minor allele decreased the probability of being self-employed.

The replication strategy for the 38 suggestive SNPs from the sex-stratified meta-analysis that were carried forward into the replication stage was similar to that used for the meta-analysis replication of the pooled data. We performed an *in silico* replication study using the data from the STR. None of the SNPs reached nominal significance ($p<0.05$) in the replication study for males only ($n=1{,}409$, Table S5) and females only ($n=1{,}862$, Table S6). In addition, for the majority of the suggestive SNPs, the direction of the effect was not consistently in the same direction as was reported in the majority of the discovery studies, again indicating that these SNPs were potential false positives. We meta-analyzed the results from the sex-stratified discovery meta-analysis and the replication study in a combined meta-analysis. For males, five

Table 3. Variance in the tendency to engage in self-employment explained by all autosomal SNPs in a combined sample of RS-I and STR for pooled males and females, males only, and females only.

Sample	σ_g^2/σ_P^2	s.e.	*p*-value	*n*	Cases	(%)	Controls	(%)
Pooled	0.25	0.14	0.032	6,223	905	(14.5)	5,318	(85.5)
Males	0.25	0.24	0.152	2,986	618	(20.7)	2,368	(79.3)
Females	0.00	0.28	0.499	3,835	353	(9.2)	3,482	(90.8)

The genetic relationships were estimated from 301,115 directly genotyped autosomal SNPs that were available in both studies. All analyses controlled for age, study, and the first 10 principal components of the genetic similarity matrix of the combined sample of RS-I and STR. In the pooled sample we also controlled for sex. The results did not change markedly when 4 or 20 principal components were included; σ_g^2/σ_P^2: proportion of phenotypic variance explained by the variance of the total additive genetic effects of the 301,115 autosomal SNPs; s.e.: standard error; *p*-value: *p*-value from a likelihood ratio (LR) test assuming that the LR is distributed as a 50:50 mixture of zero and χ_1^2.
doi:10.1371/journal.pone.0060542.t003

Table 4. Descriptive statistics of the sixteen discovery studies and the replication study.

	Pooled		Males		Females		Demographics	
Study	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**	**Mean age**	**SD age**
AGES	529	2,690	439	913	90	1,777	51.2	6.5
ASPS	46	788	26	336	20	452	65.2	8.1
ERF	214	857	113	366	101	491	47.2	13.4
GHS	424	2,706	282	1,332	142	1,374	55.9	10.9
H2000	228	1,895	145	890	83	1,005	50.7	11.1
HBCS	265	1,459	141	595	124	864	61.5	2.9
HRS	1947	4273	1048	1780	899	2493	63.6	7.9
KORA S4	177	1,588	121	760	56	828	53.8	8.8
NFBC1966	462	3,772	322	1,718	140	2,054	31.0	0.0
NTR1	201	1,354	94	494	107	860	46.4	13.3
NTR2	166	818	77	355	89	463	51.0	13.8
RS-I	531	4,843	319	1,994	212	2,849	68.8	8.8
RS-II	197	1,869	113	848	84	1,021	64.8	8.0
RS-III	209	1,716	138	746	71	970	56.1	5.8
SardiNIA	740	3,402	515	1,207	225	2,195	46.3	17.1
SHIP	157	3,906	107	1,891	50	2,015	49.7	16.3
THISEAS	204	481	176	243	28	238	51.1	11.2
TwinsUK[a]	822	2,333	–	–	730	2,165	54.5	12.4
YFS	215	2,143	89	1,194	126	949	37.6	5.0
Total discovery	7,734	42,893	4,265	17,662	3,377	25,063	53.4	9.4
STR	737	2,534	484	925	253	1,609	60.6	4.3
Total combined	8,471	45,427	4,749	18,587	3,630	26,672	53.8	9.1

AGES: Age, Gene/Environment Susceptibility–Reykjavik Study; ASPS: Austrian Stroke Prevention Study; ERF: Erasmus Rucphen Family study; GHS: Gutenberg Health Study; H2000: Health 2000; HBCS: Helsinki Birth Cohort Study; HRS: Health and Retirement Study; KORA S4: Cooperative Health Research in the Region of Augsburg; NFBC1966: Northern Finland Birth Cohort 1966; NTR1: Netherlands Twin Register Cohort 1; NTR2: Netherlands Twin Register Cohort 2; RS-I: Rotterdam Study Baseline; RS-II: Rotterdam Study Extension of Baseline; RS-III: Rotterdam Study Young; SardiNIA: SardiNIA Study of Aging; SHIP: Study of Health in Pomerania; THISEAS: The Hellenic study of Interactions between SNPs & Eating in Atherosclerosis Susceptibility; TwinsUK: the UK Adult Twin Registry; YFS: the Cardiovascular Risk in Young Finns Study; STR: Swedish Twin Registry; Cases: number of participants that were at least once self-employed; Controls: number of participants that were not, and ideally never, self-employed; SD: standard deviation.

[a]The number of male participants was insufficient for a male stratified analysis.

doi:10.1371/journal.pone.0060542.t004

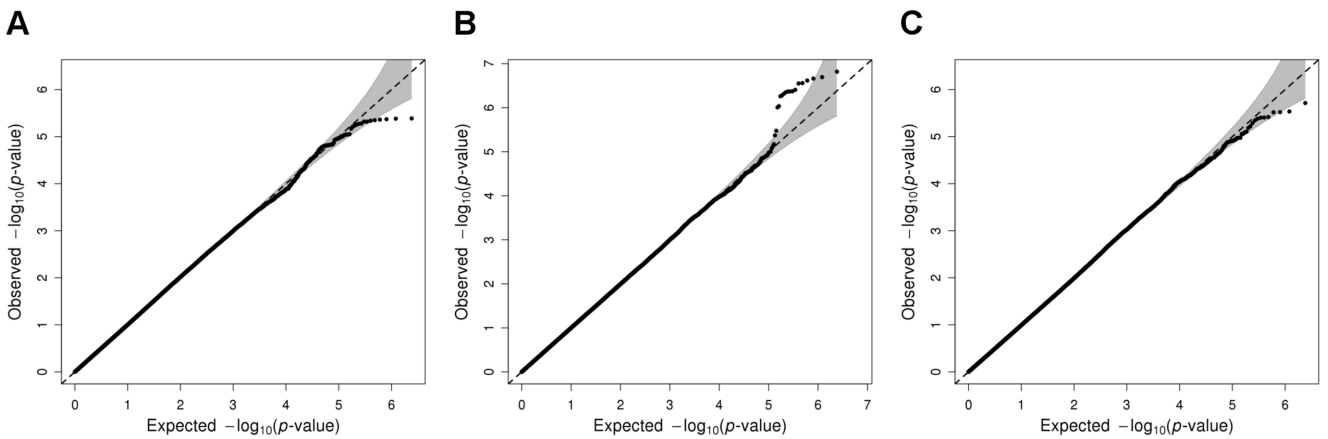

Figure 1. Q–Q plots of the self-employment discovery meta-analyses. Q–Q plot of the self-employment discovery meta-analysis for (A) pooled males and females, (B) males only, and (C) females only. The grey shaded areas in the Q–Q plots represent the 95% confidence bands around the *p*-values.

doi:10.1371/journal.pone.0060542.g001

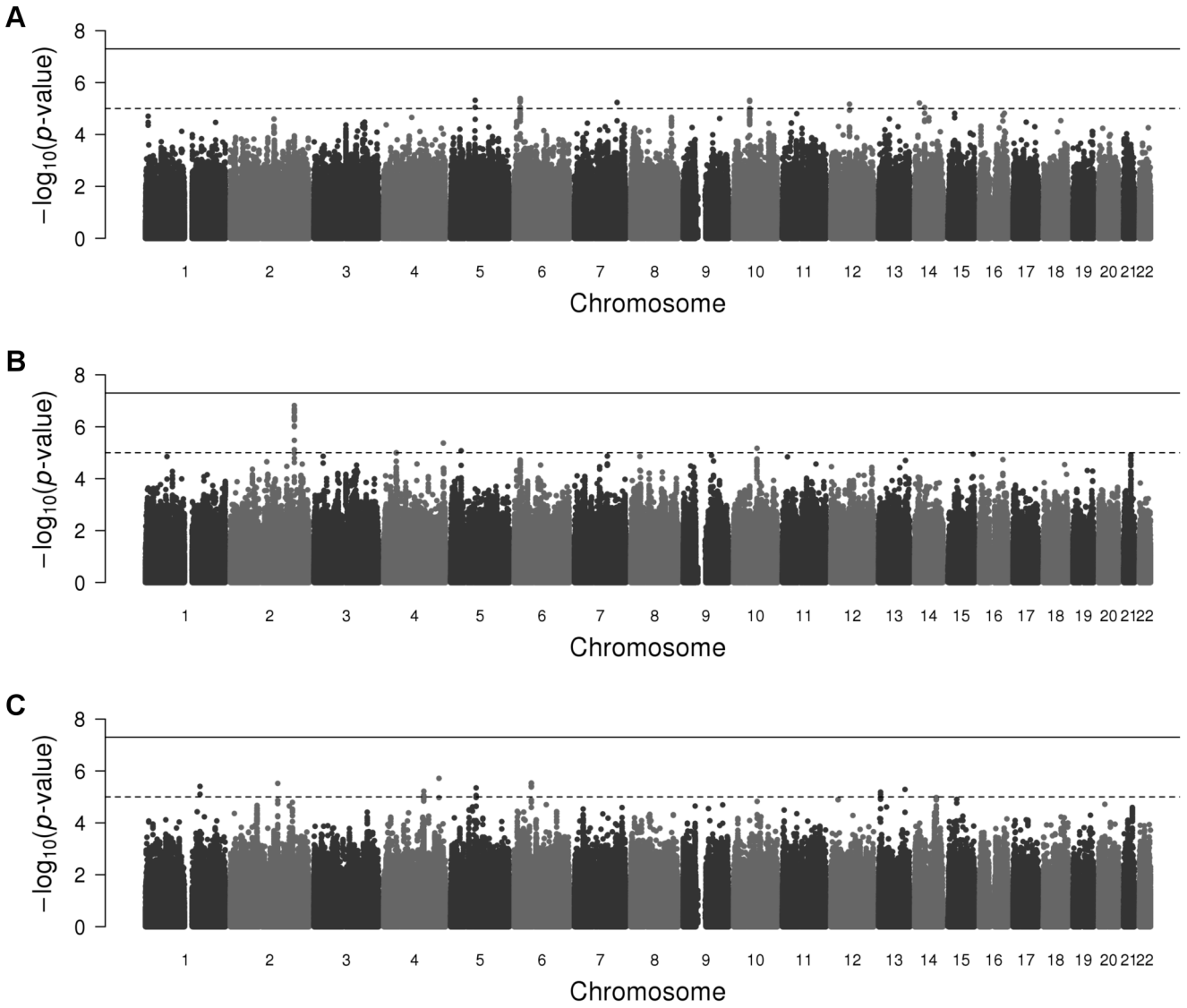

Figure 2. Manhattan plots of the self-employment discovery meta-analyses. Manhattan plot of the self-employment discovery meta-analysis for (A) pooled males and females, (B) males only, and (C) females only. SNPs are plotted on the *x*-axis according to their position on each chromosome against association with self-employment on the *y*-axis (shown as −log10 *p*-value). The solid line indicates the threshold for genome-wide significance ($p<5\times10^{-8}$) and the dashed line the threshold for suggestive SNPs ($p<1\times10^{-5}$).
doi:10.1371/journal.pone.0060542.g002

SNPs had lower *p*-values compared to the male discovery meta-analysis, although none reached genome-wide significance (Table S5). In the combined meta-analysis for females, we observed that one SNP, rs562487, had a smaller *p*-value in this combined meta-analysis; however, this SNP did not reach genome-wide significance ($p=4.01\times10^{-6}$; Table S6).

Gene-based association analyses

The findings from the discovery meta-analyses were used to perform gene-based association tests for seventeen genes that have been previously suggested to be candidate genes for entrepreneurship [48,79], including *ADORA2A*, *ADRA2A*, *COMT*, *DDC*, *DRD1*, *DRD2*, *DRD3*, *DRD4*, *DRD5*, *DYX1C1*, *HTR1B*, *HTR1E*, *HTR2A*, *KIAA0319* (*DYX2*), *ROBO1*, *SLC6A3* (*DAT1*), and *SNAP25*. Genes with $p<0.003$ (0.05/17 genes) were considered significant, but none of the candidate genes reached this level (Table S7).

To identify novel genes that may be associated with self-employment, we tested 17,697 genes for pooled males and females, 17,698 genes for males only, and 17,699 genes for females only, implying a significance level of $p<2.8\times10^{-6}$. None of the analyzed genes reached this predetermined significance level (Tables S8, S9, and S10). The gene with the lowest *p*-value was *SLC15A3* for the pooled male and female analysis ($p=1.63\times10^{-4}$). For males only, the lowest *p*-value was for *TMEM156* (1.61×10^{-4}), and for females only, the lowest *p*-value was for *PCP4* ($p=4.70\times10^{-5}$).

We also sought to replicate the association that was reported by Nicolaou et al. [48] to exist between a common variant, rs1486011, which is located in the *DRD3* gene, and the tendency to be an entrepreneur. The SNP was nominally significant in the discovery meta-analysis ($p=0.011$; Table S11); however, most studies reported a positive effect of the C allele–*opposite* to that reported by Nicolaou et al. [48], corroborating the results from an earlier replication study [49]. We also sought to replicate this SNP

Table 5. Top SNPs ($p<1\times10^{-5}$) from the self-employment discovery meta-analyses for pooled males and females, males only, and females only.

SNP	Chr.	Pos.	Effect /non-effect allele	EAF	*p*-value	Direction	Nearest gene	Number of SNPs in region
Pooled								
rs6906622	6	18,596,287	T/C	0.21	4.10×10^{-6}	++−++++++−++++?++	*RNF144B*	12
rs2358531	5	75,515,542	A/G	0.71	4.79×10^{-6}	−−−?−−−−−+−−+−−?−−	*SV2C*	2
rs10776614	10	49,433,172	T/C	0.16	4.79×10^{-6}	−+−−−−+−−−−−−−−?−−	*ARHGAP22*	2
rs17166082	7	131,363,900	A/G	0.06	5.82×10^{-6}	−?−?−−?−−+−−−−−?−−	*PLXNA4*	1
rs994208	14	33,531,622	C/G	0.66	6.11×10^{-6}	−+−−−−−−−−−−−−−−?−−	*EGLN3*	1
rs3847697	12	57,282,257	T/C	0.44	6.79×10^{-6}	−−+−−−−−?−+−+−?−?−−	*LRIG3*	1
rs3742467	14	49,709,284	T/C	0.88	9.11×10^{-6}	+++++−?+−+++−+++?++	*SOS2*	1
Males								
rs6738407	2	196,851,876	A/G	0.20	1.52×10^{-7}	−−−−−−−−−+−−−−−−?−	*HECW2*	18
rs6825440	4	183,636,063	A/T	0.24	4.25×10^{-6}	−+−−−−−−−−+−−−+−?−	*ODZ3*	1
rs7904494	10	72,056,694	A/T	0.78	6.74×10^{-6}	+−+−−−?−−++−−−−−?−	*PRF1*	1
rs4867424	5	32,331,331	T/C	0.49	8.39×10^{-6}	−−+−−−−−−−−−−−−−?−	*MTMR12*	1
rs2712008	4	38,752,396	T/G	0.14	9.94×10^{-6}	+−++++?+++++−+++?+	*KLHL5*	1
Females								
rs2331548	4	170,199,179	A/G	0.96	1.93×10^{-6}	??+?++++++++++++?++	*CBR4*	1
rs521326	6	52,927,336	A/G	0.61	2.92×10^{-6}	−−−−−−−−−−−−+−−−?−−	*GSTA4*	5
rs1022335	2	145,813,253	A/T	0.37	3.02×10^{-6}	−−−−−−?−−−−−+−−−?−−	*ZEB2*	1
rs10753804	1	168,583,032	T/C	0.49	3.92×10^{-6}	−−−−−−?−−−−+−−−−?−−	*SCYL1BP1*	2
rs562487	5	78,442,190	A/G	0.48	4.49×10^{-6}	+++++−++−+−++−++?++	*BHMT*	2
rs9557259	13	99,031,403	T/C	0.06	5.16×10^{-6}	??−?++?++++++?????+	*TM9SF2*	1
rs1383043	4	123,562,066	A/G	0.38	6.05×10^{-6}	−−+−−−−+−−−−−−−−??+	*ADAD1*	2
rs9578700	13	23,775,308	A/G	0.67	6.53×10^{-6}	−+++−−−−−−−−−−−−?−+	*SPATA13*	2

Chr.: chromosome; Pos.: position; EAF: average effect allele frequency; In the column ''direction'', the studies are in the following order: 1. AGES, 2. ASPS, 3. ERF, 4. GHS, 5. H2000, 6. HBCS, 7. HRS, 8. KORA, 9. NFBC1966, 10. NTR1, 11. NTR2, 12. RS-I, 13. RS-II, 14. RS-III, 15. SardINIA, 16. SHIP, 17. THISEAS, 18. TwinsUK (pooled and female sample)/YFS (male sample), 19. YFS (pooled and female sample); A question mark indicates that the SNP was not tested in that specific study; For SNPs that were located close together in the same region, only the most significant SNP is included in the table. The last column shows the number of neighboring SNPs that exceed the threshold for suggestive SNPs.

doi:10.1371/journal.pone.0060542.t005

in the sex-stratified discovery meta-analyses. In this analysis, we observed a certain degree of evidence for a positive effect of the C allele in males ($p = 0.046$; Table S11) but not in females ($p = 0.112$; Table S11).

Predicting self-employment from genotype data

We examined whether the results from the discovery meta-analyses could be used to predict self-employment in the replication study [76]. We pruned the set of autosomal SNPs to a subset of approximately 120,000 SNPs that are in approximate linkage equilibrium. In an initial prediction analysis, we included only the subset of these 120,000 SNPs that reached a 1% significance level. We calculated a predictive score for each individual in the replication study by determining, for each SNP, the product of the individual's number of effect alleles and the estimated regression coefficient from the discovery meta-analysis. This product was then summed across the included SNPs and divided by the number of included SNPs. We evaluated the predictive power of the SNPs by calculating the degree of variance in the tendency to engage in self-employment that was explained by the score and the area under the receiver operating characteristic curve (AUC). We repeated this prediction analysis eleven additional times, each time with a less stringent significance threshold required for a SNP to be included in the score. Hence, each time this analysis was performed, a larger subset of the 120,000 SNPs was analyzed.

For the pooled analysis of males and females ($n = 2{,}589$), the variance that was explained by the score reached a maximum of 0.184% when all SNPs were included ($p = 0.039$; Table S12). The scores for males only ($n = 1{,}110$) and for females only ($n = 1{,}479$) showed no evidence for association with self-employment (all $p \geq 0.144$, Table S12). Furthermore, we did not observe a consistent positive relationship between the variance in the tendency to engage in self-employment that was explained by the score and the significance threshold p_T (Figure 3).

Discussion

We present results from four methods of analysis, three of which are based on genome-wide molecular genetic data, to investigate the molecular genetic architecture of self-employment.

First, using a classical twin design, we report that 55% of the variance in the tendency to engage in self-employment is due to additive genetic effects, with higher heritability for males (67%) than for females (40%). Our estimates are in agreement with those of previous twin studies. These earlier studies suggested heritabilities of 48% in a sample of primarily female British twins [26] and

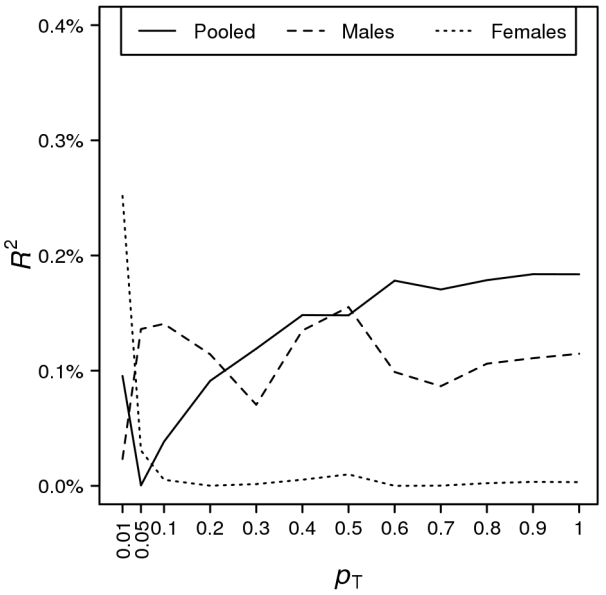

Figure 3. Prediction results. Variance explained (Nagelkerke pseudo-R^2 from logistic regression) vs. *p*-value threshold p_T for including SNPs in the score calculation.
doi:10.1371/journal.pone.0060542.g003

of 38% in a sample of US twins [28]. In addition, Zhang et al. [27] estimated the heritability of current business ownership and self-employment in a sample of Swedish twins and observed evidence of a significant additive genetic effect for females but not for males. Our results suggest significant heritability among males as well; however, the confidence intervals of the estimates are very wide for both our study and for that of Zhang et al. [27]. At least a portion of the differences between these two studies may be explained by imprecision and/or by the different samples and definitions of entrepreneurship that were used.

Second, by applying a method that was recently developed by Yang et al. [52] to entrepreneurship, we estimate that approximately 25% of the variance in the tendency to engage in self-employment (about half of the h^2 estimated in twin studies) could in principle be explained by the additive effects of common SNPs that are in linkage disequilibrium with the unknown causal variants. These results are in line with previous studies, which have estimated that common SNPs account for one-quarter to half of the narrow-sense heritability for height [52], intelligence [80,81], personality [51,82], several common diseases [83], schizophrenia [84], and recently for several economic and political preferences [22].

Several explanations may explain why the heritability estimate for self-employment using common SNPs is approximately half of the estimate that was obtained using the classical twin design. First, the causal variants may be in regions of the genome that are currently not covered by the available SNP arrays. Second, it is possible that the genotyped SNPs and the causal variants are not in complete linkage disequilibrium because, for example, the true causal variants have on average lower minor allele frequencies than the genotyped SNPs. Yang et al. [52] provide evidence for this in the case of human height. They estimated that 45% of the variance in height is accounted for by common SNPs, while the heritability of height is consistently estimated to be approximately 80%. The authors then developed a method that estimated the variance that was accounted for by common SNPs, assuming imperfect linkage disequilibrium between the genotyped SNPs and the unobserved causal variants. This method revealed that 84% of the variance in height, the complete heritability, could be explained by the causal variants. Twin and family studies do not suffer from this issue, as genetic relatedness is inferred from the expected relationships within the pedigree and include all of the additive genetic variation. Both of these explanations imply that the estimates that we obtained for self-employment using the more novel method are at the lower bounds of the heritability that is commonly estimated in twin and family studies. A third, alternative, explanation for the different results that were obtained using these techniques is that the twin-based heritability estimates are biased upwards because of, for example, genetic interactions [85] or a violation of the identical common environment assumption in twin studies [86].

Third, we perform the first meta-analysis of GWASs of an economic behavior (i.e., self-employment) using data from sixteen studies that together comprise approximately 50,000 participants. The discovery stage had 80% power to detect a variant at genome-wide significance with a minor allele frequency of 0.25 and odds ratios of approximately 1.11 for pooled males and females, 1.15 for males only, and 1.17 for females only [87], assuming we had a non-noisy, harmonized measure of self-employment across studies. Yet, we do not identify genome-wide significant associations. This result suggests that there are no common SNPs for self-employment with moderate to large effect sizes, thus placing an upper bound on the effect sizes of common SNPs that we can expect to exist. Gene-based tests for approximately 17,700 genes, including several candidate genes for entrepreneurship that have been previously suggested in the literature [48,79], do not reveal significant associations. In addition, we are unable to replicate a previously reported correlation, namely, rs1486011, a SNP that is located in the *DRD3* gene. This common variant was identified by Nicolaou et al. [48], who reported its association with the tendency to be an entrepreneur. The non-replication of associations is common in candidate gene studies of human traits and behaviors. This failure to identify replicable associations is likely due to a combination of underpowered sample sizes (due to optimistic assumptions regarding plausible effect sizes) and publication bias [88]. Examples of non-replication of candidate genes studies on complex human traits include general intelligence [81], personality [89–94], and trust [95,96]. We therefore stress that caution is warranted when interpreting claims from candidate gene studies of SNPs or genes with strong effects on complex behavioral traits like self-employment.

Finally, we report that a genetic score that was estimated in our meta-analysis sample has only limited predictive power in our replication study. The variance that was explained by the score was always lower than 0.26%. However, this result does not contradict our finding that approximately half of the narrow-sense heritability can be explained by common SNPs. This latter heritability analysis uses the measured SNPs to estimate realized relatedness between individuals, and given the large number of SNPs in a dense SNP array, realized relatedness can be estimated fairly accurately. In contrast, estimating a strongly predictive score from a sample requires good estimates of the effects of individual SNPs. If our discovery sample was infinitely large, it would have been possible to precisely estimate all of the SNP effects and to obtain a score with the theoretically highest possible predictive power, as estimated using the Yang et al. [52] method. The smaller the discovery sample, the noisier the estimates of the individual SNP effects; therefore, the predictive power of the score will be lower [97,98]. Our estimates of the effects of the individual SNPs are still too imprecise to allow out-of-sample prediction with SNP data that would have practical utility.

Together, our results demonstrate that common SNPs jointly account for a substantial share of the variance in the tendency to engage in self-employment ($\sigma_g^2/\sigma_P^2 = 25\%$). However, because we do not find specific SNPs in our large-scale meta-analyses of GWASs that examined self-employment, this heritability is not due to SNPs with moderate to large effects. A plausible interpretation of these results therefore appears to be that the molecular genetic architecture of self-employment is highly polygenic, implying that there are hundreds or thousands of variants that individually have a small effect and which together explain a substantial proportion of the heritability. We cannot rule out the possibility that rare genetic variants, or other, currently unmeasured, variants that are insufficiently correlated with the SNPs on the genotyping platforms, have large effects on an individual's tendency to be self-employed. However, if these genetic variants are rare, they would still not contribute a great deal to the population-based variance in self-employment, and large samples would still be required to identify these variants [51,83,99].

Our results are similar to those that have been reported for biologically more proximate human traits [51,52,80–82] and diseases [76,83,84] for which a polygenic molecular genetic architecture has also been suggested. One implication of this similarity is that, with sufficiently large sample sizes, SNPs that are associated with self-employment–and possibly also other economic variables–can in principle be discovered, as has been the case for, e.g., height [100] and BMI [101]. However, a discovery sample of approximately 50,000 individuals is apparently still too small for a meta-analysis of GWASs on a biologically distal, complex, and relatively rare human behavior such as self-employment. A potential opportunity for future research are GWASs of endophenotypes such as risk preferences, confidence, and independence. The effect sizes of individual SNPs on these endophenotypes may be larger because of their greater biological proximity. However, these variables are difficult to measure reliably and not (yet) available in many genotyped samples.

Given the need for very large samples in meta-analyses of GWASs on complex traits, an important challenge of the present study was to identify a measure of entrepreneurship that is available in a sufficiently large sample. We opted to maximize the available sample size in this study and operationalized entrepreneurship as self-employment, which is also the most frequently used measure of entrepreneurship in the economics literature [102].

We included every study we were aware of in the analysis that included a measure of self-employment and which was willing to contribute data, although this approach necessitated that data from diverse populations (e.g., Eastern German self-employed individuals and US business owners) were pooled. The available measures of self-employment varied across studies, including different single- and multiple-item measures, data from stand-alone surveys, and data from repeated measures or retrospective employment histories of the participants. For a number of studies, this approach resulted in a lack of detailed and reliable data regarding work-life history. Substantial measurement error, especially with respect to the definition of the control group, was therefore unavoidable. Ideally, the control group would encompass only participants who had never been self-employed and who will never be self-employed. Such an analysis would have required data regarding the complete work-life history of participants and participants who had reached an appropriate age. However, only data regarding current employment status were available in the majority of the contributing studies. It is therefore possible that there was a certain degree of misclassification in the studies that included only single-item, single-response measures of self-employment, thereby adding noise to the phenotype definition and potentially reducing the statistical power with respect to association detection.

Statistical power may have also been reduced by heterogeneity within the case group, as this group comprised individuals who became self-employed for very different reasons. For example, certain individuals may have chosen self-employment because they had no viable alternatives in paid employment, whereas others may have done so because of their desire to pursue a business opportunity. The motivations, goals, and resources of these two groups of individuals are obviously very different, and the genetics underlying these various characteristics may likewise differ greatly. Unfortunately, more detailed information regarding the motivations, activities, and success of entrepreneurs was unavailable for most of the genotyped samples.

In general, GWASs face a practical trade-off between phenotype quality and sample size. Surprisingly, statistical power calculations suggest that studying a more noisy phenotype in a larger sample is often more likely to be successful than studying a perfect phenotype in a small sample. For example, assume that a common SNP exists with a minor allele frequency of 0.5 that increases the odds for all types of entrepreneurship by a factor of 1.13 on average (assuming 15% of the population are entrepreneurs and the data are population samples). The required sample size to detect this SNP with 80% power for a perfectly-measured outcome is approximately 30,000. Measuring entrepreneurship perfectly would require a lengthier survey that is administered more than once. Such a large genotyped sample with perfect measures of entrepreneurship does not currently exist. Smaller samples with perfect measures would be underpowered to detect the SNP. In contrast, if the available measures for entrepreneurship are noisy and have a test-retest reliability of only 0.6-which is typical for behavioral traits measured by brief surveys [103–105]−80% power to detect this SNP requires a discovery sample of approximately 50,000 individuals. Thus, our study was well-powered to detect effects of this magnitude even if there was substantial measurement error and noise in the data.

The results of our study have three implications for this future research agenda. First, the high share of variance in self-employment that can be attributed towards interpersonal differences in common SNPs suggests that this research agenda is in principle feasible. Second, to investigate if and how genes that are related to economic variables influence medical outcomes, it will be necessary in the future to identify either the specific genetic variants that are underlying the heritability of economic variables (i.e., to investigate causal pathways from genes to medical outcomes), or to calculate genetic scores that have at least moderate out-of-sample predictive power (i.e., to investigate the medical consequences of a mismatch between genetic predisposition and economic outcomes). Even larger samples than what we had available in our present study will be needed to identify genome-wide significant SNPs and to estimate more accurate genetic scores for economic variables. Third, our results suggest that the effects of single SNPs on self-employment are likely to be very small. Given these effect sizes, statistical power calculations suggests that a research strategy that aims to maximize sample size by pooling data with slightly inaccurate measures of self-employment is more likely to be successful than a research strategy that aims to collect perfect phenotype measures in a much smaller sample. If successful, this research could shed new light on the complex interaction of genes, environment, and personal choices on health and longevity.

Supporting Information

Table S1 Study design, sample size, sample quality control, and self-employment measure within each study.
(DOC)

Table S2 Genotyping, imputation, SNP quality control, and statistical analysis within each study.
(DOC)

Table S3 Genomic inflation factors.
(DOC)

Table S4 Replication results of the twenty suggestive SNPs ($p<1\times10-5$) from the self-employment discovery meta-analyses for pooled males and females.
(DOC)

Table S5 Replication results of the 22 suggestive SNPs ($p<1\times10-5$) from the self-employment discovery meta-analyses for males only.
(DOC)

Table S6 Replication results of the sixteen suggestive SNPs ($p<1\times10-5$) from the self-employment discovery meta-analyses for females only.
(DOC)

Table S7 Gene-based p-values for the candidate entrepreneurship genes for pooled males and females, males only, and females only.
(DOC)

Table S8 Gene-based p-values for the top 25 genes associated with self-employment in the discovery meta-analysis for pooled males and females.
(DOC)

Table S9 Gene-based p-values for the top 25 genes associated with self-employment in the discovery meta-analysis for males only.
(DOC)

Table S10 Gene-based p-values for the top 25 genes associated with self-employment in the discovery meta-analysis for females only.
(DOC)

Table S11 Meta-analysis association results for SNP rs1486011 for pooled males and females, males only, and females only.
(DOC)

Table S12 Results of the prediction analyses in STR for pooled males and females, males only, and females only.
(DOC)

Acknowledgments

We are grateful to Peter Visscher for his helpful comments and suggestions; AGES: The researchers are indebted to the participants for their willingness to participate in the study; ASPS: The authors thank the staff and the participants of the ASPS for their valuable contributions. We thank Birgit Reinhart for her long-term administrative commitment and Ing Johann Semmler for technical assistance with the creation of the DNA bank; ERF: We are grateful to all of the patients and their relatives, as well as to the general practitioners and the neurologists for their contributions. We are also thankful to P. Veraart for her assistance in matters pertaining to genealogy, Jeannette Vergeer for supervision of the laboratory work, and P. Snijders for his assistance in data collection; GHS: We thank all of the study participants and all of the colleagues that are involved in the GHS; H2000: We would like to thank all of the Health 2000 Survey participants; HBCS: We thank all of the study participants as well as everyone who is involved in the Helsinki Birth Cohort Study; NFBC1966: We thank Professor Paula Rantakallio (launch of NFBC1966 and initial data collection), Ms. Sarianna Vaara (data collection), Ms. Tuula Ylitalo (administration), Mr. Markku Koiranen (data management), Ms. Outi Tornwall and Ms. Minttu Jussila (DNA biobanking); NTR: We thank all of the participating twin families for their cooperation; RS: We thank Pascal Arp, Mila Jhamai, Dr. Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their assistance in creating the GWAS database. The authors are very grateful to the participants and staff from the Rotterdam Study, the participating general practitioners and the pharmacists. We would also like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf, as well as their institutions: the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and ; Services@MediGRID part of the German D-Grid for access to their grid resources; SardiNIA: We thank all of the volunteers who participated in the study, Monsignore Piseddu, Bishop of Ogliastra, the mayors and citizens of the participating Sardinian towns (Lanusei, Ilbono, Arzana, and Elini), the head of the Public Health Unit ASL4 for their volunteer work and cooperation, and the team of biologists, physicians, nurses, and the recruitment personnel; SHIP: The contributions to data collection made by the field workers, the study physicians, the ultrasound technicians, the interviewers, and the computer assistants are gratefully acknowledged; STR: The STR thanks the SNP&SEQ Technology Platform, Uppsala for their genotyping assistance; THISEAS: The Hellenic study of Interactions between SNPs and Eating in Atherosclerosis Susceptibility (THISEAS) thanks the genotyping facility at the Wellcome Trust Sanger Institute for typing the THISEAS samples and, in particular, Sarah Edkins and Cordelia Langford. We also thank all of the dieticians and clinicians for their contribution to the project; TwinsUK: We would like to thank the TwinsUK twins for their continuing support and participation in our studies. We thank Dr. Lynn Cherkas for her involvement in this work; YFS: Irina Lisinen and Ville Aalto are gratefully acknowledged for their expert technical assistance in the statistical analyses.

Author Contributions

Individual study design and management: GRA RB SB DIB DC FC EJCdG GD PD GE JE CG VG AH M-RJ MJ LJL BAO MP SR DS R. Schmidt COS TDS AT CMvD HV H-EW PSW GW. Data collection: SB DIB DC EJCdG MD JE RH M-RJ MJ M. Kähönen JL TL PKEM KP OR R. Schmidt AVS TDS FJAvR JV PSW GW. Genotyping: SB PD J-JH TL PKEM FR HS AVS AGU PSW GW. Genotype preparation: GRA J-JH TL PKEM FR HS AS AVS AGU. Phenotype preparation: GAA-B SEB SB DC NE BH M. Kähönen JL ML SN KP OR AVS AT MJHMvdL FJAvR JV PSW GW. Study data analysis: GRA SEB NE J-JH AI M. Kaakinen M. Kähönen SK MAL TL ML KP OR CAR AS PS AVS IS ET MJHMvdL JV SMW. Manuscript review: SEB DJB SB DIB DC GD GE NE PJFG AH RH J-JH MJ PDK MAL PKEM MP LQ FR DS R. Schmidt COS AS AVS R. Svento AT ART AGU FJAvR HV PSW. Analysis plan development: PDK MJHMvdL. Analysis plan review: FR FJAvR AGU. Meta-analyses: NE MJHMvdL. Manuscript preparation: CAR MJHMvdL. Heritability, accounted-for variance by common SNPs, and prediction analyses: CAR MJHMvdL. Review and interpretation of analyses: PJFG AH PDK CAR FR ART AGU MJHMvdL FJAvR. Conceived and designed the study: AH PDK ART. Organize and oversee consortium: AH ART.

References

1. Marmot MG, Kogevinas M, Elston MA (1987) Social/economic status and disease. Annu Rev Public Health 8: 111–135.
2. Adler NE, Boyce T, Chesney MA, Cohen S, Folkman S, et al. (1994) Socioeconomic status and health: The challenge of the gradient. Am Psychol 49: 15–24.

3. Adler NE, Ostrove JM (1999) Socioeconomic status and health: What we know and what we don't. Ann N Y Acad Sci 896: 3–15.
4. Steenland K, Henley J, Thun M (2002) All-cause and cause-specific death rates by educational status for two million people in two American cancer society cohorts, 1959–1996. Am J Epidemiol 156: 11–21.
5. Van Kippersluis JLW, O'Donnell OA, van Doorslaer EKA (2011) Long run returns to education: Does education lead to an extended old age? J Hum Resour 94: 695–721.
6. Lager ACJ, Torssander J (2012) Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. Proc Natl Acad Sci USA 109: 8461–8466.
7. Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR (1989) Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. Am J Epidemiol 129: 1132–1144.
8. Winkleby MA, Jatulis DE, Frank E, Fortmann SP (1992) Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. Am J Public Health 82: 816–820.
9. Ettner SL (1996) New evidence on the relationship between income and health. J Health Econ 15: 67–85.
10. Dowd JB, Albright J, Raghunathan TE, Schoeni RF, LeClere F, et al. (2011) Deeper and wider: Income and mortality in the USA over three decades. Int J Epidemiol 40: 183–188.
11. Kaplan GA, Keil JE (1993) Socioeconomic factors and cardiovascular disease: A review of the literature. Circulation 88: 1973–1998.
12. Haynes SG, Feinleib M (1980) Women, work and coronary heart disease: Prospective findings from the Framingham Heart Study. Am J Public Health 70: 133–141.
13. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747–753.
14. McGue M, Vaupel JW, Holm N, Harvald B (1993) Longevity is moderately heritable in a sample of Danish twins born 1870–1880. J Gerontol 48: B237–B244.
15. Herskind AM, McGue M, Holm NV, Sørensen TI, Harvald B, et al. (1996) The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. Hum Genet 97: 319–323.
16. Mitchell BD, Hsueh WC, King TM, Pollin TI, Sorkin J, et al. (2001) Heritability of life span in the Old Order Amish. Am J Med Genet 102: 346–352.
17. VB Hjelmborg J, Iachine I, Skytthe A, Vaupel JW, McGue M, et al. (2006) Genetic influence on human lifespan and longevity. Hum Genet 119: 312–321.
18. Behrman J, Taubman P (1976) Intergenerational transmission of income and wealth. Am Econ Rev 66: 436–440.
19. Miller P, Mulvey C, Martin N (2001) Genetic and environmental contributions to educational attainment in Australia. Econ Educ Rev 20: 211–224.
20. Scarr S, Weinberg RA (1994) Educational and occupational achievements of brothers and sisters in adoptive and biologically related families. Behav Genet 24: 301–325.
21. Lichtenstein P, Pedersen NL, McClearn G (1992) The origins of individual differences in occupational status and educational level. Acta Sociol 35: 13–31.
22. Benjamin DJ, Cesarini D, van der Loos MJHM, Dawes CT, Koellinger PD, et al. (2012) The molecular genetic architecture of economic and political preferences. P Natl Acad Sci USA 109: 8026–8031.
23. Björklund A, Jäntti M, Solon G (2007) Nature and nurture in the intergenerational transmission of socioeconomic status: Evidence from Swedish children and their biological and rearing parents. BE J Econ Anal Poli 7(2): article 4.
24. Sacerdote B (2007) How large are the effects from changes in family environment? A study of Korean American adoptees. Q J Econ 122: 119–157.
25. Taubman P (1976) The determinants of earnings: Genetics, family, and other environments: A study of white male twins. Am Econ Rev 66: 858–870.
26. Nicolaou N, Shane S, Cherkas L, Hunkin J, Spector TD (2008) Is the tendency to engage in entrepreneurship genetic? Manage Sci 54: 167–179.
27. Zhang Z, Zyphur MJ, Narayanan J, Arvey RD, Chaturvedi S, et al. (2009) The genetic basis of entrepreneurship: Effects of gender and personality. Organ Behav Hum Dec 110: 93–107.
28. Nicolaou N, Shane S (2010) Entrepreneurship and occupational choice: Genetic and environmental influences. J Econ Behav Organ 76: 3–14.
29. Cooper CL, Marshall J (1976) Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. J Occup Psychol 49: 11–28.
30. Cooper CL, Smith M (1985) Job Stress and Blue Collar Work. Chichester, UK: Wiley.
31. Argyle M (1997) Is happiness a cause of health? Psychol Health 12: 769–781.
32. Schnall PL, Landsbergis PA, Baker D (1994) Job strain and cardiovascular disease. Annu Rev Public Health 15: 381–411.
33. Beauchamp JP, Cesarini D, Johannesson M, van der Loos MJHM, Koellinger PD, et al. (2011) Molecular genetics and economics. J Econ Perspect 25: 57–82.
34. Benjamin DJ, Cesarini D, Chabris CF, Glaeser EL, Laibson DI, et al. (2012) The promises and pitfalls of genoeconomics. Annu Rev Econ 4: 627–662.
35. Lewin-Epstein N, Yuchtman-Yaar E (1991) Health risks of self-employment. Work Occupation 18: 291–312.
36. Dahl MS, Nielsen J, Mojtabai R (2010) The effects of becoming an entrepreneur on the use of psychotropics among entrepreneurs and their spouses. Scand J Public Health 38: 857–863.
37. Hamilton BH (2000) Does entrepreneurship pay? An empirical analysis of the returns to self-employment. J Polit Econ 108: 604–631.
38. Blanchflower DG, Oswald AJ (1998) What makes an entrepreneur? J Labor Econ 16: 26–60.
39. Block J, Koellinger PD (2009) I can't get no satisfaction–necessity entrepreneurship and procedural utility. Kyklos 62: 191–209.
40. Benz M, Frey BS (2008) Being independent is a great thing: Subjective evaluations of self-employment and hierarchy. Economica 75: 362–383.
41. Shane S, Venkataraman S (2000) The promise of entrepreneurship as a field of research. Acad Manage Rev 25: 217–226.
42. Andersson L, Hammarstedt M (2010) Intergenerational transmissions in immigrant self-employment: Evidence from three generations. Small Bus Econ 34: 261–276.
43. Colombier N, Masclet D (2008) Intergenerational correlation in self employment: Some further evidence from French ECHP data. Small Bus Econ 30: 423–437.
44. Dunn T, Holtz-Eakin D (2000) Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. J Labor Econ 18: 282–305.
45. Evans DS, Leighton LS (1989) Some empirical aspects of entrepreneurship. Am Econ Rev 79: 519–535.
46. Lentz BF, Laband DN (1990) Entrepreneurial success and occupational inheritance among proprietors. Can J Economics 23: 563–579.
47. Van der Zwan PW, Thurik AR, Grilo I (2010) The entrepreneurial ladder and its determinants. Appl Econ 42: 2183–2191.
48. Nicolaou N, Shane S, Adi G, Mangino M, Harris J (2011) A polymorphism associated with entrepreneurship: Evidence from dopamine receptor candidate genes. Small Bus Econ 36: 151–155.
49. Van der Loos MJHM, Koellinger PD, Groenen PJ, Rietveld CA, Rivadeneira F, et al. (2011) Candidate gene studies and the quest for the entrepreneurial gene. Small Bus Econ 37: 269–275.
50. Visscher PM, Goddard ME, Derks EM, Wray NR (2012) Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. Mol Psychiatry 17: 474–485.
51. Verweij KJ, Yang J, Lahti J, Veijola J, Hintsanen M, et al. (2012) Maintenance of genetic variation in human personality: Testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. Evolution 66: 3238–3251.
52. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565–569.
53. Koellinger PD, van der Loos MJHM, Groenen PJF, Thurik AR, Rivadeneira F, et al. (2010) Genome-wide association studies in economics and entrepreneurship research: Promises and limitations. Small Bus Econ 35: 1–18.
54. Van der Loos MJHM, Koellinger PD, Groenen PJF, Thurik AR (2010) Genome-wide association studies and the genetics of entrepreneurship. Eur J Epidemiol 25: 1–3.
55. Du Rietz A, Henrekson M (2000) Testing the female underperformance hypothesis. Small Bus Econ 14: 1–10.
56. Bird B, Brush C (2002) A gendered perspective on organizational creation. Entrep Theory Pract 26: 41–65.
57. Georgellis Y, Wall HJ (2005) Gender differences in self-employment. Int Rev Appl Econ 19: 321–342.
58. Koellinger P, Minniti M, Schade C (2011) Gender differences in entrepreneurial propensity. Oxford B Econ Stat: In press.
59. Verheul I, Thurik A, Grilo I, van der Zwan P (2012) Explaining preferences and actual involvement in self-employment: Gender and the entrepreneurial personality. J Econ Psychol 33: 325–341.
60. Riding AL, Swift CS (1990) Women business owners and terms of credit: Some empirical findings of the Canadian experience. J Bus Venturing 5: 327–340.
61. Verheul I, Thurik AR (2001) Start-up capital: "Does gender matter?". Small Bus Econ 16: 329–346.
62. Bates T (2002) Restricted access to markets characterizes women-owned businesses. J Bus Venturing 17: 313–324.
63. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, et al. (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. Circ Cardiovasc Genet 2: 73–80.
64. Neale MC, Boker SM, Xie G, Maes HH (2003) Mx: Statistical modeling. Richmond, VA: Virginia Commonwealth University, Department of Psychiatry.
65. Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet 88: 76–82.
66. Dempster ER, Lerner IM (1950) Heritability of threshold characters. Genetics 35: 212–236.
67. So HC, Li M, Sham PC (2011) Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet Epidemiol 35: 447–456.
68. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.
69. Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55: 997–1004.

70. Willer CJ, Li Y, Abecasis GR (2010) METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26: 2190–2191.
71. Higgins JPT, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21: 1539–1558.
72. Higgins JPT, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327: 557–560.
73. Cochran WG (1954) The combination of estimates from different experiments. Biometrics 10: 101–129.
74. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101.
75. Liu JZ, Mcrae AF, Nyholt DR, Medland SE, Wray NR, et al. (2010) A versatile gene-based test for genome-wide association studies. Am J Hum Genet 87: 139–145.
76. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460: 748–752.
77. Falconer DS (1960) Introduction to quantitative genetics. New York: Ronald Press.
78. Pearson TA, Manolio TA (2008) How to interpret a genome-wide association study. JAMA 299: 1335–1344.
79. Shane S (2010) Born entrepreneurs, born leaders: How your genes affect your work life. New York: Oxford University Press.
80. Davies G, Tenesa A, Payton A, Yang J, Harris SE, et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatr 16: 996–1005.
81. Chabris CF, Hebert BM, Benjamin DJ, Beauchamp JP, Cesarini D, et al. (2012) Most reported genetic associations with general intelligence are probably false positives, Psychol Sci 23: 1314–1323.
82. Vinkhuyzen AAE, Pedersen NL, Yang J, Lee SH, Magnusson PKE, et al. (2012) Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. Transl Psychiatry 2: e102.
83. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet 88: 294–305.
84. Lee SH, Decandia TR, Ripke S, Yang J, Sullivan PF, et al. (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nat Genet 44: 247–250.
85. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci USA 109: 1193–1198.
86. Charney E (2008) Genes and ideologies. Perspect Polit 6: 299–319.
87. Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: Design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19: 149–150.
88. Ioannidis JPA (2005) Why most published research findings are false. PLOS Med 2(8): e124.
89. Ebstein RP, Novick O, Umansky R, Priel B, Osher Y, et al. (1996) Dopamine D4 receptor (*D4DR*) exon III polymorphism associated with the human personality trait of novelty seeking. Nat Genet 12: 78–80.
90. Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, et al. (1996) Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. Science 274: 1527–1531.
91. Paterson AD, Sunohara GA, Kennedy JL (1999) Dopamine D4 receptor gene: Novelty or nonsense? Neuropsychopharmacol 21: 3–16.
92. Terracciano A, Balaci L, Thayer J, Scally M, Kokinos S, et al. (2009) Variants of the serotonin transporter gene and NEO-PI-R Neuroticism: No association in the BLSA and SardiNIA samples. Am J Med Genet B Neuropsychiatr Genet 150B: 1070–1077.
93. Verweij KJH, Zietsch BP, Medland SE, Gordon SD, Benyamin B, et al. (2010) A genome-wide association study of Cloninger's temperament scales: Implications for the evolutionary genetics of personality. Biol Psychol 85: 306–317.
94. De Moor MHM, Costa PT, Terracciano A, Krueger RF, de Geus EJC, et al. (2012) Meta-analysis of genome-wide association studies for personality. Mol Psychiatr 17:337–349.
95. Israel S, Lerer E, Shalev I, Uzefovsky F, Riebold M, et al. (2009) The oxytocin receptor (*OXTR*) contributes to prosocial fund allocations in the dictator game and the social value orientations task. PLOS ONE 4(5): e5535.
96. Apicella CL, Cesarini D, Johannesson M, Dawes CT, Lichtenstein P, et al. (2010) No association between oxytocin receptor (*OXTR*) gene polymorphisms and experimentally elicited social preferences. PLOS ONE 5(6): e11153.
97. Goddard ME, Wray NR, Verbyla K, Visscher PM (2009) Estimating effects and making predictions from genome-wide marker data. Statist Sci 24: 517–529.
98. Visscher PM, Yang J, Goddard ME (2010) A commentary on 'Common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010) Twin Res Hum Genet 13: 517–524.
99. Wray NR, Purcell SM, Visscher PM (2011) Synthetic associations created by rare variants do not explain most GWAS results. PLOS Biol 9(1): e1000579.
100. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832–838.
101. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 42: 937–948.
102. Parker SC (2009) The economics of entrepreneurship. Cambridge, UK: Cambridge University Press.
103. Loomis JB (1989) Test-retest reliability of the contingent valuation method: A comparison of general population and visitor responses. Am J Agr Econ 71: 76–84.
104. Weertman A, Arntz A, Dreessen L, van Velzen C, Vertommen S (2003) Short-interval test-retest interrater reliability of the Dutch version of the Structured Clinical Interview for DSM-IV personality disorders (SCID-II). J Pers Disord 17: 562–567.
105. Ansolabehere S, Rodden J, Snyder JM (2008) The strength of issues: Using multiple measures to gauge preference stability, ideological constraint, and issue voting. Am Polit Sci Rev 102: 215–232.